41 | 从感知机到神经网络算法

2019-01-31 李智慧

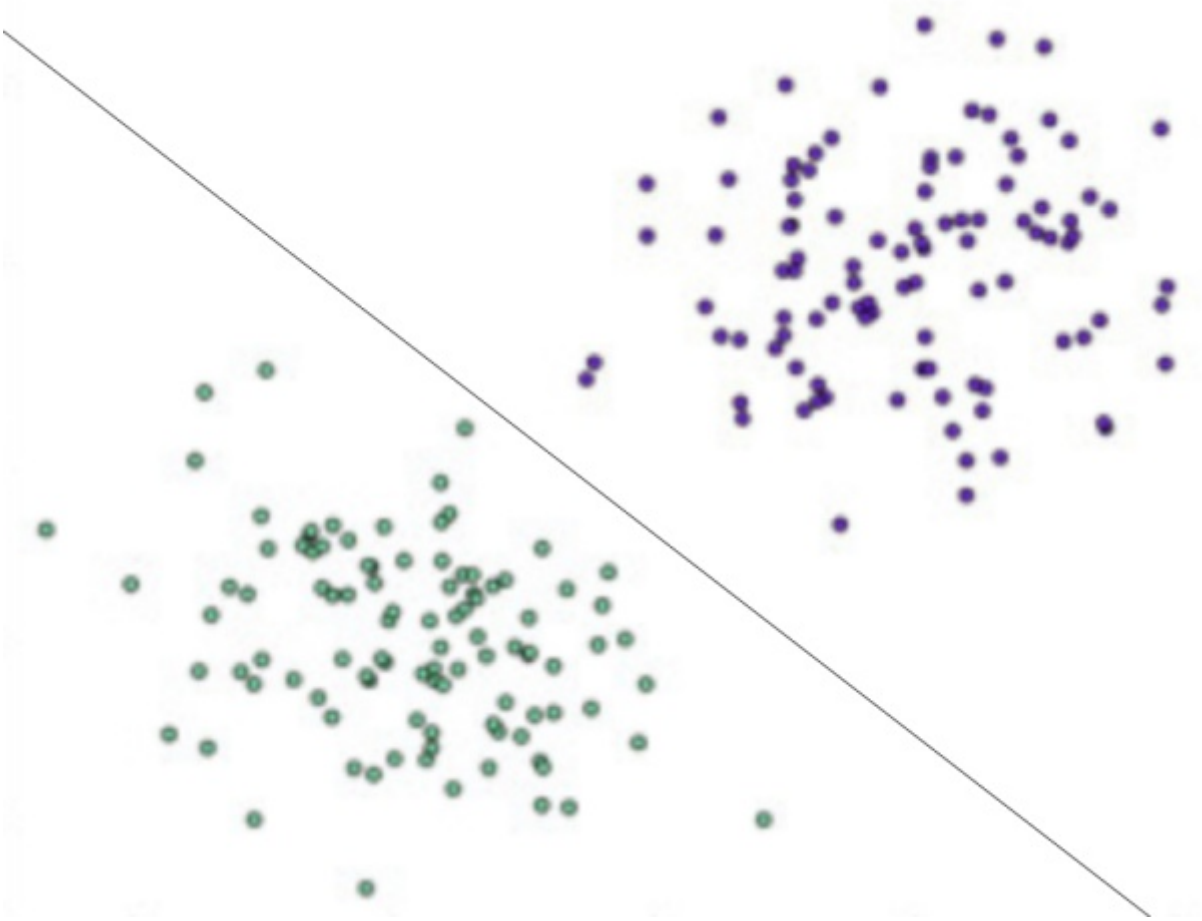
从0开始学大数据 进入课程 >

讲述：李智慧
时长 10:00 大小 9.18M

从机器学习模型角度看，目前最简单的机器学习模型大概就是感知机了，而最火热的机器学习模型则是神经网络。人工智能领域几乎所有炫酷的东西都是神经网络的成果，有下赢人类最顶尖围棋棋手的 AlphaGo、自动驾驶技术、聊天机器人、语音识别与自动翻译等。事实上，神经网络和感知机是一脉相承的，就像复杂的人体是由一个个细胞组成、复杂的大脑是由一个个神经元组成，而神经网络正是由感知机组成的。

感知机

感知机是一种比较简单的二分类模型，将输入特征分类为 +1、-1 两类，就像下图所示的，一条直线将平面上的两类点分类。



二维平面上的点只有两个输入特征（横轴坐标和纵轴坐标），一条直线就可以分类。如果输入数据有更多维度的特征，那么就需要建立同样多维度的模型，高维度上的分类模型也被称为超平面。

感知机模型如下：

$$f(x) = sign(w \cdot x + b)$$

其中 x 代表输入的特征空间向量，输出空间是{-1, +1}，w 为权值向量，b 叫作偏置，$sign$ 是一个符号函数。

$$\text{sign}(x) = \begin{cases} +1, & x \geqslant 0 \\ -1, & x < 0 \end{cases}$$

$w \cdot x + b = 0$ 为超平面的方程，当感知机输出为 +1 表示输入值在超平面的上方，当感知机输出为 -1 表示输入值在超平面的下方。训练感知机模型就是要计算出 w 和 b 的值，当有新的数据需要分类的时候，输入感知机模型就可以计算出 +1 或者 -1 从而进行分类。

由于输出空间只有{-1, +1}两个值，所以只有误分类的时候，才会有模型计算值和样本真实值之间的偏差，偏差之和就是感知机的损失函数。

$$L(w,b) = -\sum_{x_i \in M} y_i(w \cdot x_i + b)$$

其中 M 为误分类点集合，误分类点越少，损失函数的值越小；如果没有误分类点，损失函数值为 0。求模型的参数 w 和 b，就是求损失函数的极小值。

数学上求函数的极小值就是求函数的一阶导数，但是感知机损失函数用统计求和函数表达，没办法计算解析解。机器学习采用梯度下降法求损失函数极小值，实质上就是求导过程的数值计算方法。

对于误分类点集合 M，损失函数 $L(w,b)$ 变化的梯度，就是某个函数变量的变化引起的函数值的变化，根据感知机损失函数可知：

$$\Delta_w L(w,b) = -\sum_{x_i \in M} y_i x_i$$

$$\Delta_b L(w,b) = -\sum_{x_i \in M} y_i$$

使用梯度下降更新 w 和 b，不断迭代使损失函数 $L(w,b)$ 不断减小，直到为 0，也就是没有误分类点。感知机算法的实现过程：

1. 选择初始值 w_0, b_0。

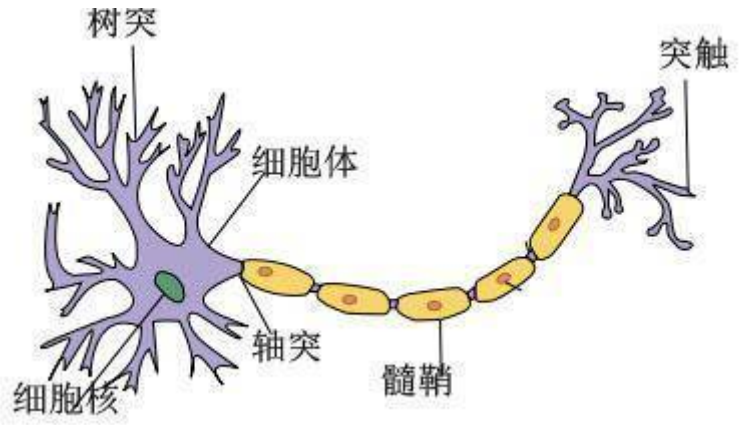
2. 在样本集合中选择样本数据 x_i, y_i。

3. 如果 $y_i(w \cdot x_i + b) < 0$，表示 y_i 为误分类点，那么 $w = w + \eta y_i x_i$、$b = b + \eta y_i$，在梯度方向校正 w 和 b。其中 η 为步长，步长选择要适当，步长太长会导致每次计算调整太大出现震荡；步长太短又会导致收敛速度慢、计算时间长。

4. 跳转回 2，直到样本集合中没有误分类点， 即全部样本数据 $y_i(w \cdot x_i + b) \geq 0$。

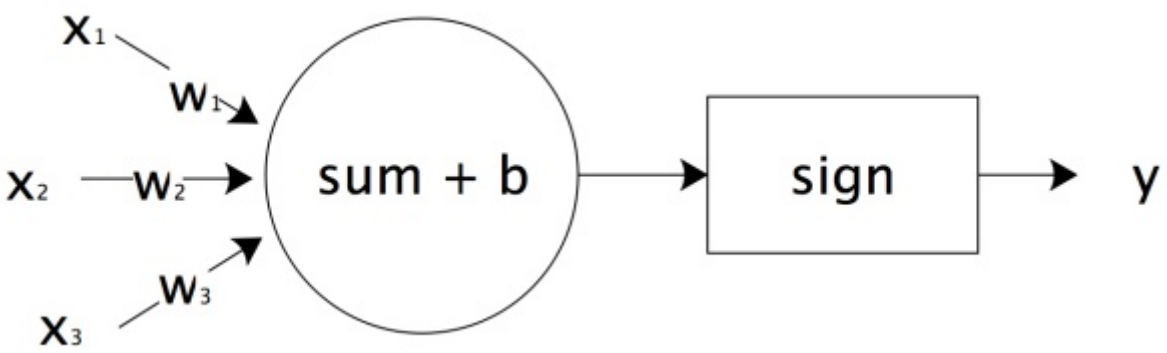
神经网络

我们现在所说的神经网络，通常是指机器学习所使用的“人工神经网络”，是对人脑神经网络的一种模拟。人脑神经网络由许多神经元构成，每个神经元有多个树突，负责接收其他神经元的输出信号，神经元细胞完成对输入信号的处理，转换成输出信号，通过突触传递给其他神经元。



神经元细胞的输出只有 0 或者 1 两种输出，但是人脑大约有 140 亿个神经元，这些神经元组成一个神经网络，前面的神经元输出作为后面的神经元输入进一步处理，最终实现人类的智能。

人脑神经元可以通过感知机进行模拟，每个感知机相当于一个神经元，使用 $sign$ 函数的感知机输出也是只有两个值，跟人脑神经元一样。

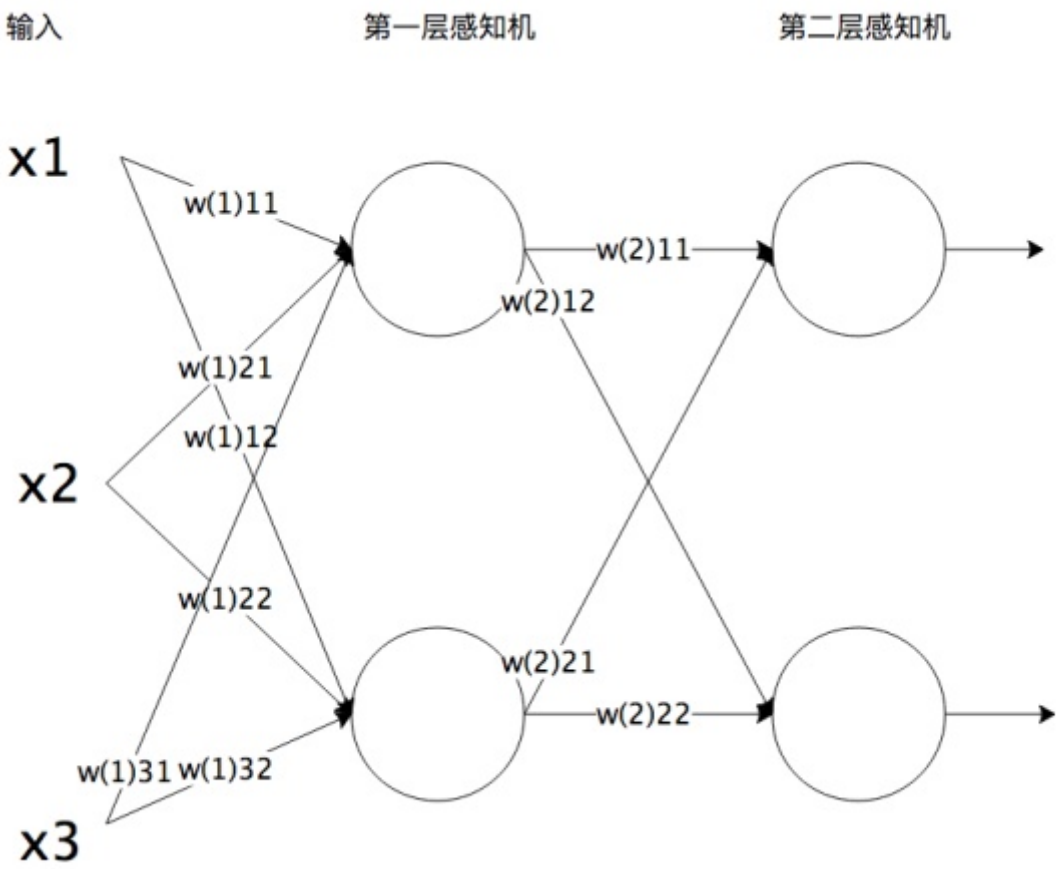


x_1, x_2, x_3 相当于神经元的树突，实现信号的输入；$sum() + b$ 及 $sign$ 函数相当于神经元细胞，完成输入的计算；y 是神经元的输出，上图用数学形式表达的话是

$$y = sign(w_1x_1 + w_2x_2 + w_3x_3 + b)$$

它是感知机 $y = sign(w \cdot x + b)$ 向量展开形式。

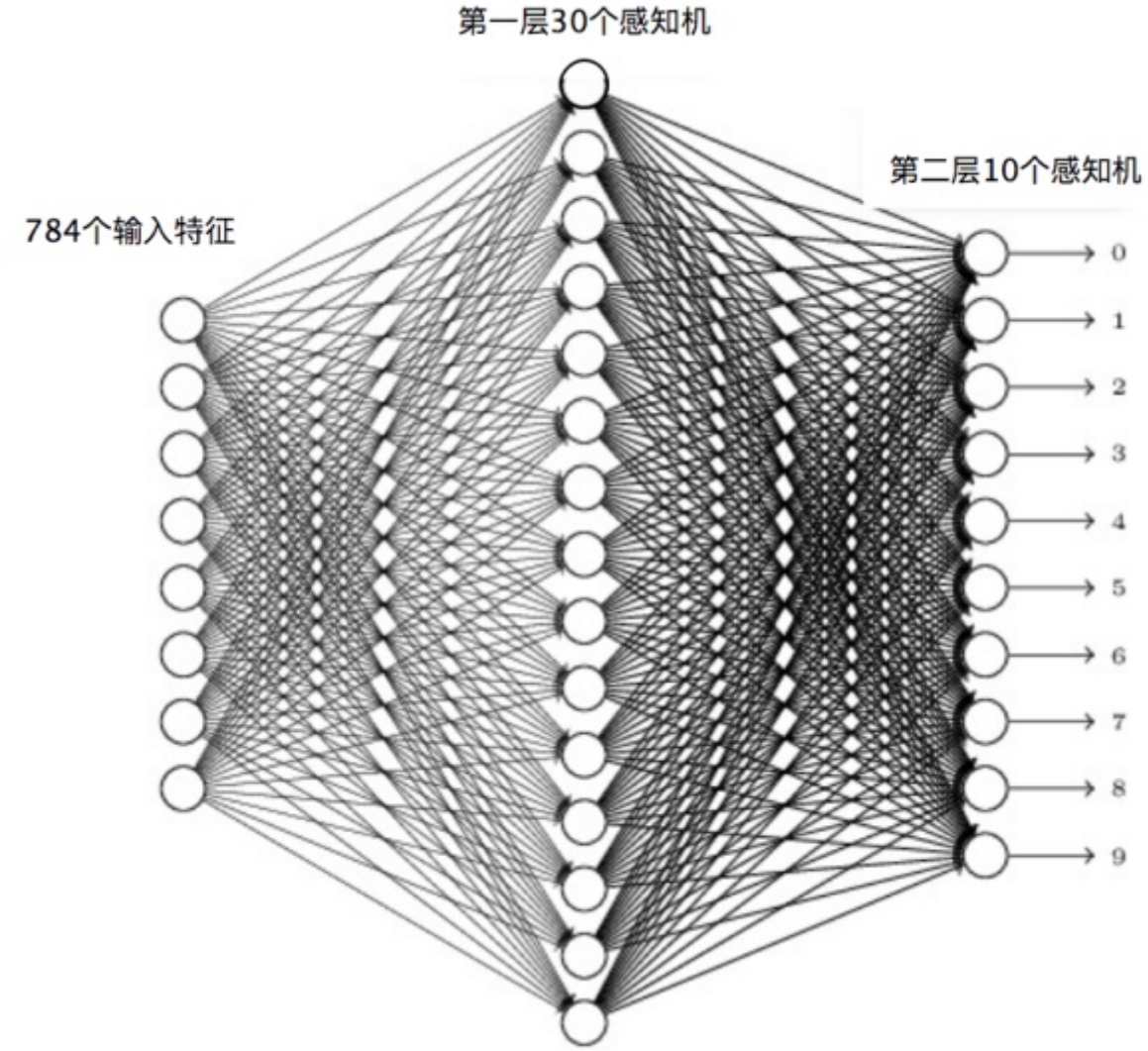
将感知机组成一层或者多层网络状结构，就构成了机器学习神经网络。下图就是一个两层神经网络。



在多层神经网络中，每一层都由多个感知机组成。将输入的特征向量 x 传递给第一层的每一个感知机，运算以后作为输出传递给下一层的每一个感知机，直到最后一层感知机产生最终的输出结果。这就是机器学习神经网络的实现过程，通过模拟人脑神经网络，利用样本数据训练每个感知机神经元的参数，在某些场景下得到的模型可以具有不可思议的效果。

以神经网络实现手写数字识别为例，样本如下。

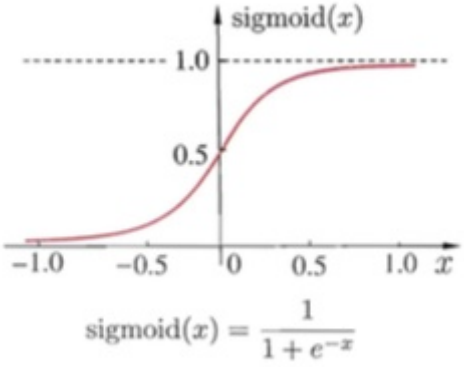
这个手写数字样本中的每个数字都是一个 28×28 像素的图片，我们把每个像素当作一个特征值，这样每个数字就对应 784 个输入特征。因为输出需要判别 10 个数字，所以第二层（输出层）的感知机个数就是 10 个，每个感知机通过 0 或者 1 输出是否为对应的数字。



使用梯度下降算法，利用样本数据，可以训练神经网络识别手写数字，计算每个感知机的 w 和 b 参数值。当所有的感知机参数都计算出来，神经网络也就训练出来了。这样对于新输入的手写数字图片，可以进行自动识别，输出对应的数字。

训练神经网络的时候采用一种反向传播的算法，针对每个样本，从最后一层，也就是输出层开始，利用样本结果使用梯度下降算法计算每个感知机的参数。然后以这些参数计算出来的结果作为倒数第二层的输出计算该层的参数。然后逐层倒推，反向传播，计算完所有感知机的参数。

当选择两层神经网络的时候，原始感知机的 $sign$ 函数表现并不太好，更常用的是 $sigmoid$ 函数。

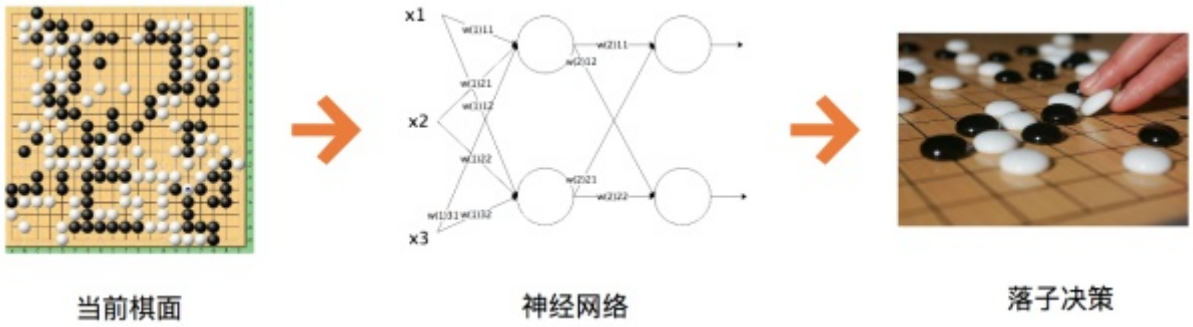


对于两层以上的多层神经网络，$ReLU$ 函数的效果更好一些。$ReLU$ 函数表达式非常简单

$$y = max(x, 0)$$

当 x 大于 0，输出 x；当 x 小于 0，输出 0。

神经网络根据组织和训练方式的不同有很多类型。当神经网络层数比较多的时候，我们称它们为深度学习神经网络。前两年在人工智能领域大放异彩的围棋程序 AlphaGo 则是一种卷积神经网络。



对于一个 19×19 的围棋棋盘，在下棋过程中，每个位置有黑、白、空三种状态，将其提取为特征就是神经网络的输入（事实上，输入特征还需要包括气、眼、吃等围棋规则盘面信息）。而输出设置 19×19 即 361 个感知机产生对应的落子。然后将大量人类的棋谱，即当前盘面下的最佳落子策略作为训练样本，就可以训练出一个智能下棋的神经网络。

但是这样根据人类棋谱训练得到神经网络最多就是人类顶尖高手的水平，AlphaGo 之所以能够碾压人类棋手还依赖一种叫蒙特卡洛搜索树的的算法，对每一次落子以后的对弈过程进行搜索，判断出真正的最佳落子策略。利用蒙特卡洛搜索树结合神经网络，AlphaGo 还可以进行自我对弈，不断进行自我强化，找到近乎绝对意义上的最优落子策略。

小结

神经网络的应用目前在大数据领域越来越广泛，很多传统机器学习模型的算法逐步尝试用神经网络代替。一般说来，传统的机器学习算法的结果是可以解释的，KNN 算法的分类结果为什么是这样，贝叶斯分类的结果为什么是这样，都是可以利用样本数据和算法来解释的。如果分类效果不好，是样本数据有问题，还是算法过程有问题，也都可以分析出来。但是一般认为，神经网络计算的结果是不可解释的，为什么神经网络会分类输出这样的结果，人们无法解释；输出结果不满意，也无法找到原因，只能不断尝试。

神经网络中每个感知机的参数可以通过训练获得，也就是 w 和 b 可以计算得到，但是一个神经网络应该设置多少层，每层应该有多少个感知机神经元，这些参数必须要算法工程师设置，因此这些参数也被称为超级参数。超级参数如何设置目前还没有太好的方法，只能依赖算法工程师的经验和不断尝试去优化。

思考题

你认为强人工智能是否会出现呢？人类有一天会被机器人统治吗？

欢迎你点击“请朋友读”，把今天的文章分享给好友。也欢迎你写下自己的思考或疑问，与我和其他同学一起讨论。

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇　40 | 机器学习的数学原理是什么?

下一篇　42 | 模块答疑：软件工程师如何进入人工智能领域?

精选留言 (12)

写留言

Hyun
2019-01-31

2

机器能否取代人类，首要条件是机器必须拥有人类基本的能力，其次是机器必须要有政治意识。

人类拥有的智慧之一就是劳动能力，包括制造机器人。那么机器人只要能改造机器人就成。但培养政治意识就难了。机器人百毒不侵，不没有感知神经，不会受到人类弱点的...

展开

miketan
2019-05-08

强人工智能能否出现，需要有标准去验证。对于人工智能否统治人类，个人觉得这是另外一个问题，表达意思就是即便到不到强人工智能也有可能统治人类。就像人类历史上野蛮落后也能战胜先进文明。

展开

Liber
2019-03-28

感知机为什么表现的不好？一开始没理解，看看这个可以当做补充：
https://www.jianshu.com/p/e4c1686ca4ed

梦归幽
2019-03-19

之前在课堂上学深度学习神经网络课程时听的一知半解迷迷糊糊的，但是这次在老师的专栏里看老师再说一遍感觉就醍醐灌顶茅塞顿开了，老师的讲解真的很读到，很适合有基础

的人再进一步理解

展开

暴风雪

2019-02-24

老师，我不明白数字识别的第一层感知机个数为什么是30个？

展开

暴风雪

2019-02-23

我觉得自己的数学基础不错了，没想到。。。。

展开

张苗

2019-02-07

强人工智能就是认为有可能制造出真正能推理和解决问题的智能机器，并且，这样的机器能将被认为是有知觉的，有自我意识的。
我认为在现阶段以及未来20年内都不会出现强人工智能。现在的人工智能技术无论是监督学习还是非监督学习，当然神经网络也属于监督学习，都是以数据和算法为基础的，而数据决定了算法效果的上限。这些导致强人工智能在现有的技术基础上无法实现。除非在...

展开

杰之7

2019-02-03

通过这一节的阅读学习，了解了从感知机到神经网络的实现原理。感知机通过输入特征训练感知机模型f(x)=sign(wx+b)的w，b参数。当出现误分类时，通过梯度下降降维求解极小值，也就是w，b值。对于误分类的集合，通过调整步长，使yi(w+xi+b)>=0
,老师，对于大于等于零时我的一个疑问是越接近零是不是效果越好?
...

展开

蜗牛行天下

2019-01-31

关于人类是否会被机器人统治，我看到这样一种解释，比较认可。首先，结论是不可能。

原因是，机器人从本质上讲，可以看作硅基生命。地球生物在进化的过程中，必然曾进行过各种尝试，但最终的结果是所有生物基本都属于碳基生命。碳在化学稳定性上比硅更突出，所以才成为所有生物的生命基础。机器人归根到底是构建在以硅为基础的芯片上面的，是不可能在能量利用效率上超过真正的生物体。而在更有突破性的智慧领域，我觉...

展开

Twogou27

2019-01-31

老师，想问一下神经网络实现手写数字例子中，为什么第一层是30个感知机，不能直接一层10个感知机，然后利用梯度下降法来训练模型吗?

作者回复: 也能，但是效果不如神经网络好

无形

2019-01-31

机器没有意识、欲望，不可能统治人类

展开

王亚南

2019-01-31

人工智能的算法简单来说就是找经验，先利用大量的样本数据分析，然后利用分析形成的经验来判断。人类最初的知识也是通过经验得来的，但是人可以跟进一步，从经验总结出规律来。基本过程是根据经验做出模型假设，然后用数据来验证假设，如果能够通过验证，则规律成立，否则进一步调整模型。所以，在我看来，人类比人工智能高级之处，就在于能够做出假设，等到那天人工智能也可以自己做出假设并完成验证，强人工智能也...

展开